



LC3: un modèle spatial et sémantique pour découvrir la connaissance dans les jeux de données géospatiaux

Benjamin Harbelot, Helbert Arenas, Christophe Cruz

► To cite this version:

Benjamin Harbelot, Helbert Arenas, Christophe Cruz. LC3: un modèle spatial et sémantique pour découvrir la connaissance dans les jeux de données géospatiaux. Extraction et Gestion de la Connaissance, Jan 2015, Luxembourg, Luxembourg. hal-01207393

HAL Id: hal-01207393

<https://hal.science/hal-01207393>

Submitted on 30 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LC3: un modèle spatial et sémantique pour découvrir la connaissance dans les jeux de données géospatiaux

Benjamin Harbelot*, Helbert Arenas**
Christophe Cruz ***

*Laboratoire LE2I, UMR-6306 CNRS, Checksem, Université de Bourgogne
9, rue Alain Savary, 21078 Dijon, France
benjamin.harbelot@checksem.fr,
<http://www.checksem.fr>
**helbert.arenas@checksem.fr
***christophe.cruz@u-bourgogne.fr

Résumé. Il est nécessaire pour les gérants de territoires d'avoir un aperçu des connaissances actuelles et de l'évolution de certaines caractéristiques de la biosphère. Les outils de télédétection enregistrent une grande quantité d'informations liées à la couverture terrestre permettant l'étude de processus dynamiques. Toutefois, le volume des jeux de données nécessite de nouveaux outils pour identifier des motifs et extraire de la connaissance. Nous proposons un modèle capable de découvrir la connaissance sur des données parcellaires et permettant l'analyse des phénomènes dynamiques à l'aide de données temporelles, spatiales et thématiques. Le modèle est appelé Land Cover Change Continuum (LC3) et se base sur les technologies du Web sémantique pour permettre une représentation accrue du contexte de l'environnement géospatial et fournir des résultats d'analyses proches de ceux des experts du domaine via des opérations de raisonnement automatique. En définitive, ce modèle permet d'améliorer notre compréhension de la dynamique des territoires.

1 Introduction

L'intérêt des systèmes d'information géographique pour les organisations privées et les gouvernements est croissant. Les systèmes d'information géographique (SIG) impliquent d'énormes quantités de micro et macro-données spatio-temporelles. L'un des objectifs de développement majeur concernant les futurs SIG est de fournir une analyse de l'évolution de l'environnement géospatial. Cependant, au regard des outils actuellement disponibles, il est difficile de fournir une analyse complète permettant de comprendre pleinement les dynamiques spatio-temporelles. Dans les SIG classiques, l'analyse consiste à lier ensemble les données pour leur donner un sens. Par conséquent, l'étude de la dynamique spatio-temporelle est fréquemment réduite à une analyse statistique. Une grande majorité de ces outils s'appuie sur des bases de données relationnelles, ce qui limite fortement les possibilités de prendre en compte le contexte de l'environnement géographique. La modélisation, l'analyse et la visualisation de phénomènes

dynamiques géospatiales a été identifié comme un enjeu majeur du développement des SIG de la prochaine génération. Un domaine d'étude pertinent dans le cadre de la dynamique spatiale concerne l'évolution de la couverture et de l'utilisation des terres plus connue sous l'acronyme "LULCC" (Land Use Land Cover Change). Au cours de milliers d'années d'existence, les humains ont modifié l'environnement. Toutefois, ce n'est que récemment que les scientifiques ont identifié une relation entre le LULCC et la modification des conditions météorologiques. Il existe plusieurs modèles actuellement utilisés pour modéliser le LULCC, cependant il y a un besoin permanent pour de nouvelles approches afin de réévaluer les modèles actuels et les améliorer.

Dans cet article, nous présenterons le domaine de la modélisation spatio-temporelle, puis nous établirons un comparatif entre le web sémantique et la base de données relationnelle comme support d'information. La section 3 présentera notre modèle défini à l'aide des outils du web sémantique. Enfin, nous détaillerons un exemple d'utilisation de notre modèle appliqué à l'étude de la dynamique du territoire Girondin.

2 Modélisation spatio-temporelle

Afin d'introduire le domaine d'étude, nous présenterons dans cette section quelques notions spatiales et temporelles.

2.1 Représenter les entités à travers le temps

Une entité spatio-temporelle est une représentation des entités du monde réel composées d'une identité, de propriétés descriptives et de propriétés spatiales. Bien que l'identité décrit une composante fixe de l'entité, les propriétés alphanumériques et spatiales peuvent varier au fil du temps et représentent la partie dynamique de l'entité. Lorsque l'identité d'une entité varie, un type particulier d'évolution intervient qui transforme l'entité spatio-temporelle en une nouvelle entité. Dans la littérature, il existe deux principaux types d'entités spatio-temporelles : 1) les objets en mouvement, comme par exemple un taxi se déplaçant dans les rues d'une ville, et 2) les objets changeant, par exemple, une région dont les frontières administratives évoluent dans le temps.

2.2 L'identité

Un concept important en ce qui concerne l'évolution des entités est l'identité. Elle caractérise l'unicité d'un objet indépendamment de ses attributs ou valeurs. Elle est la caractéristique qui distingue un objet de tous les autres. L'identité est essentielle dans la conception et la modélisation d'un phénomène. Son importance lors de la modélisation de systèmes dynamiques a été identifiée par des recherches antérieures comme (Del Mondo et al., 2013), (Muller, 2002). Cependant, cette notion est très subjective car elle dépend des critères choisis par l'utilisateur pour définir l'identité d'une entité. En général, les critères pour la définition de l'identité dépendent du domaine d'étude.

2.3 Les relations de filiation

Pour pallier cette limitation, les relations de filiation définissent les liens de succession qui existent entre les différentes représentations d'un même objet à des instants différents. Dès lors, cette approche suppose une représentation détaillée de la dynamique spatio-temporelle de l'environnement. Les relations de filiation sont particulièrement adaptées à la modélisation de motifs génériques de changement comme les divisions ou les fusions d'entités. En outre, ces changements spatiaux peuvent révéler des changements dans la nature de l'entité. Pour cette raison, les relations de filiation sont intimement liées à la notion d'identité. Cette relation est essentielle pour maintenir l'identité d'une entité qui évolue et pour suivre son évolution dans le temps.

Des recherches antérieures ont identifié deux grands types de relations de filiation : la continuation et la dérivation (Harbelot et al., 2014), (Stell et al., 2011). Dans le premier cas, l'identité ne varie pas. L'entité continue d'exister, mais a subi un changement. Alors que dans le second cas, une nouvelle entité est créée à partir de l'entité "parent" lors du changement subi par l'entité. Les relations de dérivation peuvent impliquer plusieurs entités en même temps.

2.4 Étude des différentes modélisations

L'évolution d'une entité spatiale dans le temps peut être vue soit comme une succession d'états (ou représentations) de l'entité, soit comme une succession de transition intervenant sur cette entité au cours du temps. Les modèles de la première proposition se basent sur des approches continues ou discrètes. Des exemples de modèles sont le modèle snapshot (Armstrong, 1988), le modèle Space-Time Composites (STC) (Langran et Chrisman, 1988), le modèle Spatio-temporal Object (Worboys, 1994). L'inconvénient de ces modèles est qu'ils ne représentent que des changements soudains au travers desquels il est difficile d'identifier des processus tels que le changement ou le mouvement d'une entité de l'environnement géographique. D'autres modèles basés sur les approches continues ou discrètes ont été proposés pour gérer les changements liés à l'identité (Hornsby et Egenhofer, 2000) ou encore les changements sur les relations topologiques étudiées à l'aide de matrices d'intersection Egenhofer et Al-Taha (1992), toutefois l'analyse des causes du changement nécessaire à l'étude des phénomènes sont difficiles à déduire en utilisant ce type de modélisation. Par conséquent, cette première approche de modélisation ne permet pas une analyse complète de l'évolution. Pour pallier à ce problème, l'approche de modélisation basée sur les événements et processus a peu à peu vu le jour. Cette approche considère que les entités spatiales évoluent sous l'impulsion d'un événement ou d'un processus et dont l'objectif est d'analyser les causes et les conséquences. Parmi ces modèles, il est possible de citer le modèle Event-Based Spatiotemporal Data Model (ESTDM) (Peuquet et Duan, 1995), les processus composites (Claramunt et Theriault, 1996) ou encore le modèle de changement topologique basé sur les événements (Jiang et Worboys, 2009). Le modèle ESTDM décrit un phénomène au travers d'une liste d'événements, un nouvel événement est créé en bout de liste à chaque fois qu'un changement est détecté. Toutefois ce modèle ne prend en compte que des données de type raster et les liens de causalité entre les événements sont difficilement mis en évidence dans ce modèle. Pour répondre à ce problème, le modèle des processus composites a pour objectif de représenter les liens entre les événements ainsi que leurs conséquences, en outre, l'auteur avance que le modèle de données doit différencier ce qui est spatial, temporel et thématique. Le modèle des processus composites se

base sur un langage permettant de décrire la sémantique liée à un phénomène du monde réel. Ce phénomène est assimilé à un processus composite, c'est-à-dire une suite de processus qui décrivent la dynamique du phénomène. Un processus composite est, par exemple, la trajectoire d'un bateau et peut se décomposer en 3 processus : stabilité, déplacement, rotation.

3 Le web sémantique comme support de l'information géographique

L'objectif d'un système d'information est de stocker des données d'une manière organisée pour modéliser un domaine. Lors de la modélisation d'un domaine d'application, deux visions différentes peuvent être appliquées pour représenter la connaissance. La première est l'hypothèse du monde fermé, plus connue sous l'acronyme CWA (Closed World Assumption), et suppose que toute déclaration qui est vraie est également connue pour être vraie. Par conséquent, ce qui n'est pas actuellement connu comme étant vrai, est faux. Le contraire de l'hypothèse du monde fermé est l'hypothèse du monde ouvert ou OWA (Open World Assumption), indiquant que le manque de connaissances ne signifie pas la fausseté. Chacune de ces hypothèses est étroitement liée à des technologies spécifiques. OWA est souvent liée au Web sémantique tandis que CWA est traditionnellement associé aux bases de données relationnelles. A l'heure actuelle, beaucoup de modèles spatio-temporels utilisent des bases de données relationnelles. Dans ces travaux, nous avons souhaité étudier le Web sémantique comme support de la modélisation spatio-temporelle. Plus spécifiquement, nous baserons notre modèle sur une ontologie à l'aide du langage OWL ¹.

3.1 Comparatif entre l'hypothèse du monde ouvert et du monde fermé

L'hypothèse du monde fermé invite à définir « ce qui est possible ». A l'inverse, l'hypothèse du monde ouvert permet de statuer « ce qui n'est pas possible ». Lorsqu'une ontologie OWL est vide alors tout est possible. Ce n'est que lorsque l'on contraint progressivement l'ontologie qu'elle devient plus restrictive. Dans les domaines bien définis (réservation de vols ou de livres), le modèle relationnel constitue une approche adaptée. L'hypothèse du monde fermé est performante pour faciliter la validation des données pendant les opérations de transaction. Le nombre de faits négatifs sur un domaine donné est généralement beaucoup plus grand que le nombre de faits positifs. Ainsi, dans de nombreuses applications, le nombre de faits négatifs est si grand que leur représentation explicite peut devenir pratiquement impossible. Dans de tels cas, il est plus simple et plus rapide de définir seulement tous les faits « vrais » connus plutôt que d'énumérer également les « faux ». Cependant, le modèle relationnel est un paradigme où l'information doit être complète et décrite par un schéma unique. Le modèle relationnel suppose que tous les objets et les relations qui existent dans le domaine sont ceux qui sont explicitement représentés dans la base de données, et qui identifient de manière unique les noms d'objets dans ce domaine. Cela rend l'hypothèse du monde fermé et ses hypothèses connexes un mauvais choix lorsque l'on tente de combiner des informations provenant de sources multiples, pour faire face à l'incertitude ou l'incomplétude du monde.

1. <http://www.w3.org/2001/sw/wiki/OWL>

3.2 Discussion

L'avantage principal de l'hypothèse du monde ouvert et du Web Sémantique (généralement associé à cette hypothèse) est de permettre aux informations d'être réutilisables. La réutilisation d'une ontologie permet d'assembler, étendre, spécialiser ou encore adapter les connaissances définies à partir d'autres ontologies. De cette manière, le Web Sémantique offre une bonne flexibilité pour permettre l'intégration de nouvelles connaissances lorsque l'application nécessite des connaissances spécifiques. Lorsqu'une ontologie est étendue, toutes les déclarations définies comme étant vraie le reste. Enfin, l'approche du Web Sémantique permet de fournir des mécanismes d'inférences afin de générer de la connaissance au sein d'une application. De son côté, l'approche relationnelle s'avère être un candidat tout indiqué pour la validation des données. Cependant, les irrégularités et l'incomplétude sont une limite à la conception du modèle relationnel. L'approche Web Sémantique dépasse ces limites en proposant une structure flexible du schéma des données. De plus, la dissociation explicite entre le schéma (TBox) et les données (ABox) offre un environnement propice pour l'interprétation des données représentées. Par conséquent, l'incomplétude du monde ouvert peut être partiellement comblée grâce à des raisonneurs capables d'auto-alimenter le système en se basant sur des contraintes. Dans beaucoup d'application du web sémantique, il est intéressant d'utiliser OWL pour définir des contraintes d'intégrités qui doivent être satisfaites par les données. Cependant, l'hypothèse du monde ouvert et le rejet de l'hypothèse du nom unique sont jusqu'alors un frein au développement des contraintes d'intégrités en OWL. En effet, les conditions définies pour lever des violations de contraintes en monde fermé, génèrent de nouvelles connaissances dans les applications de raisonnement basées sur OWL. Dès lors, il apparaît nécessaire de fournir des applications hybrides combinant à la fois les raisonnements en monde ouvert avec les capacités de validation de contraintes du monde fermé.

4 Le modèle LC3

Le modèle LC3 a été proposé pour répondre au besoin de représenter les changements qui s'opèrent sur des entités spatiales. Afin de formaliser notre modèle, nous utiliserons la logique de description (Baader et Nutt (2003)).

4.1 Modélisation d'une couche spatio-temporelle

Notre modèle tente de représenter des entités dynamiques évoluant dans le temps. Ces entités sont appelées des *timeslices* dans le cadre de nos travaux. Chacun d'entre eux se définit selon quatre composantes que sont : l'identité, l'espace, le temps, et la sémantique intrinsèque de l'entité. L'identité est la composante la plus importante du modèle. Généralement, lorsque la modélisation porte sur des parcelles de terrain, une classe correspond à une couverture du sol spécifique et définit l'identité du *timeslice*, cependant d'autres critères peuvent être utilisés pour souligner l'unicité d'un *timeslice*. Les ontologies possèdent la particularité de pouvoir organiser des classes sur différents niveaux sémantiques en utilisant une taxonomie. Chaque classe décrit un concept et la taxonomie permet d'associer les *timeslices* à des concepts plus ou moins spécifiques.

$$Timeslice \sqsubseteq \top \quad (1)$$

Dans notre modèle, la classe *Timeslice* désigne le concept plus général et peut être spécialisée par une hiérarchie. Dès lors, les concepts spécifiques sont utiles pour distinguer les entités représentées tandis que les concepts généraux permettent au contraire de les regrouper sémantiquement. Un tel agencement des concepts crée un indice de profondeur au sein de la hiérarchie et permet d'évaluer les différents niveaux sémantiques qu'il est possible d'exploiter. L'équation suivante formalise la hiérarchie.

$$C_{i+2} \sqsubseteq C_{i+1} \sqsubseteq C_i \sqsubseteq \dots \sqsubseteq C_0 \quad (2)$$

où C_0 correspond à la classe *Timeslice* et $i + 2$ représente la profondeur de la hiérarchie.

Afin de représenter le temps, nous nous basons sur l'approche suggérée par (Artale et Franconi, 1998) en considérant le domaine temporel comme une structure linéaire composée par un ensemble de points temporels (*TemporalPoint*).

$$TemporalPoint \sqsubseteq \top \quad (3)$$

Tous les éléments de type *TemporalPoint* suivent un ordre strict, qui oblige tous les points entre deux instants temporels t_1 et t_2 à être ordonné dans le temps. En sélectionnant une paire de points temporels $[t_o, t_f]$, il est possible de définir intervalle fermé de points ordonnés définissant ainsi des intervalles de temps (*Intervalle*).

$$\begin{aligned} Interval &\sqsubseteq \top \\ Interval &\equiv \exists hasStartPoint.TemporalPoint \sqcap \\ &\quad \exists hasEndPoint.TemporalPoint \end{aligned} \quad (4)$$

Dans certaines applications, la structure du temps peut également être définie selon des instants de temps (*TemporalPoint*). Afin de représenter à la fois des intervalles et des instants de temps, nous définissons le concept *Time* (\mathcal{T}).

$$Time \equiv TemporalPoint \sqcup Interval \quad (5)$$

Dans notre modèle, nous définissons la propriété *hasTime* qui possède comme *Domain* la classe *Time*, ainsi il est possible d'utiliser les classes *TemporalPoints* et *Intervals* selon les besoins de modélisation :

$$\forall hasTime.\mathcal{T} \quad (6)$$

En définitive, dans nos travaux, les différents états des entités spatio-temporelles sont représentés par la classe *TimeSlice* (\mathcal{TS}). Cette classe comprend quatre composantes : 1) spatiale,

qui est la représentation géométrique de l'entité (\mathcal{G}) ; 2) identité, qui associe chaque état à l'entité (\mathcal{O}) qu'il représente ; 3) temporelle, pour décrire le temps (\mathcal{T}) durant lequel le *timeslice* est valide ; 4) un ensemble de propriétés alphanumériques, qui décrivent les caractéristiques de l'entité durant la période de validité du *timeslice*. L'équation 7 représente la formalisation de la classe *TimeSlice* dans le cadre de nos travaux.

$$\begin{aligned} \mathcal{TS} \equiv & \exists hasGeometry.\mathcal{G} \sqcap \\ & \exists TimeSliceOf.\mathcal{O} \sqcap \\ & \exists hasTime.\mathcal{T} \sqcap \\ & \exists hasProperties.\overline{\mathcal{TS}} \end{aligned} \quad (7)$$

4.2 Modélisation d'une transition spatio-temporelle

Sur une zone géographique possédant une couverture dynamique du territoire, une même région peut être associée à différents *timeslices* entre deux instants de temps consécutifs. Afin de représenter ces associations dans le temps, nous définissons la relation *hasFiliation* dans notre modèle. Par conséquent, la classe *Timeslice* est définie par un *domain* (voir Equation 8) et un *range* (voir Equation 9).

$$\exists hasFiliation \sqsubseteq \mathcal{TS} \quad (8)$$

$$\top \sqsubseteq \forall hasFiliation.\mathcal{TS} \quad (9)$$

Cette propriété est essentielle pour représenter le lien entre deux entités. Cependant, la connaissance portée par cette propriété est trop pauvre pour bien comprendre l'évolution. Dans notre modèle, nous proposons de spécialiser cette propriété au travers de différentes couches de connaissance. Les équations 10, 11 et 12 formalisent la hiérarchie de connaissance établie sur la base de cette propriété en utilisant le principe de subsomption.

$$hasFiliation \equiv hasContinuation \sqcup hasDerivation \quad (10)$$

Afin de spécialiser nos relations de filiation, nous intégrons les contraintes d'identité pour distinguer les objets qui ont changé leur nature de ceux dont l'identité est restée invariante. Ainsi, nous distinguons deux types de filiations que sont les continuations et les dérivations.

$$\begin{aligned} & hasContinuation \equiv \\ hasEquality \sqcup hasGrowth \sqcup hasReduction \sqcup hasAnnexation \sqcup hasSeparation \end{aligned} \quad (11)$$

$$\begin{aligned} & hasDerivation \equiv \\ Conversion \sqcup hasSplit \sqcup hasFusion \sqcup partOfAnnexation \sqcup partOfSeparation \end{aligned} \quad (12)$$

Un modèle sémantique pour l'étude de la dynamique des territoires

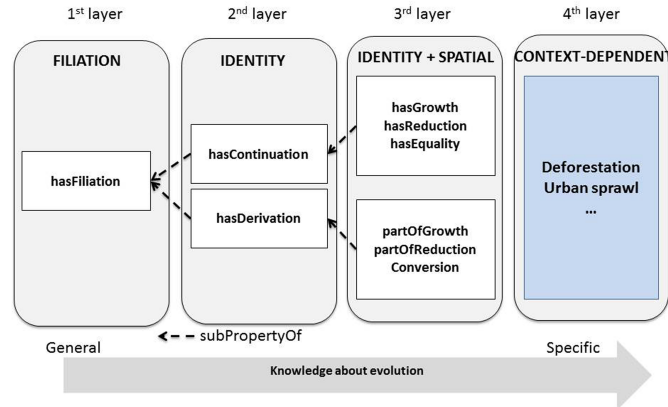


FIG. 1 – Les différentes couches pour qualifier l'évolution

La dernière couche générique de notre modèle ajoute des contraintes spatiales pour aboutir à des motifs détectables au sein du jeu de données. Ces motifs sont définis au travers des relations suivantes :

- **Expansion** : l'entité continue d'exister mais la géométrie s'étend. Cette relation est une relation de continuation.
- **Contraction** : l'entité continue d'exister mais la géométrie se réduit. Cette relation est une relation de continuation.
- **Division** : l'entité parent cesse d'exister, sa géométrie se sépare en deux nouvelles géométries correspondants chacune à une nouvelle entité qui n'existait pas auparavant. Cette relation donne lieu à deux dérivations.
- **Séparation** : l'entité parent continue d'exister mais sa géométrie se sépare et une nouvelle géométrie correspondants à une nouvelle entité apparait. Cette relation donne lieu à une continuation et au moins une dérivation.
- **Fusion** : les deux entités parents fusionnent et cessent d'exister pour donner lieu à une nouvelle géométrie correspondant à une nouvelle entité. Cette relation donne lieu à deux dérivations.
- **Annexion** : les deux entités parents fusionnent mais une seule continue d'exister. Cette relation donne lieu à une continuation et au moins une dérivation.
- **Stabilité** : aucun changement n'a eut lieu sur l'entité.
- **Conversion** : la géométrie est restée identique tandis que l'identité a varié.

5 Exemple de la gestion des risques d'inondation dans la région de la Gironde

La section précédente présente les bases de notre modèle dans l'optique de fournir un support d'analyse pour l'étude de la dynamique des territoires. Notre méthode tente de décrire et

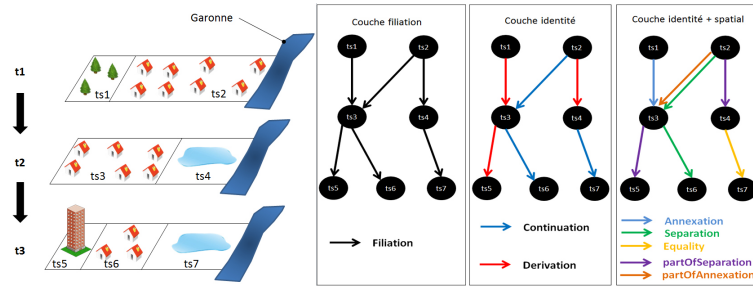


FIG. 2 – Liste des motifs génériques détecté dans le modèle LC3

d'offrir une meilleure compréhension de l'évolution appliquée à l'étude du territoire Girondin. Cet exemple se base sur le jeu de données Corine Land Cover qui offre une couverture de tout le territoire européen pour les années 1990, 2000 et 2006. Pour faciliter la compréhension, nous illustrerons notre approche en schématisant le territoire étudié. La Figure 2 présente notre exemple d'évolution ainsi que sa représentation sous forme de graphe au travers des différentes couches proposées par notre modèle. La dernière couche du modèle est la plus expressive. Elle permet de définir des motifs génériques d'évolution indépendants de tout domaine d'application pour servir de support à l'inférence de phénomènes. En revanche, les phénomènes peuvent être très nombreux. Par conséquent, il est impossible de les représenter au sein d'une couche générique car chacun d'entre eux est spécifique au domaine d'application. Pour contourner cette limitation, notre approche utilise les données contextuelles présentes au sein de l'ontologie que l'on appelle communément "ontologie de domaine". L'utilisation de cette ontologie de domaine couplée aux motifs génériques définis dans la dernière couche du modèle permet de fournir une interprétation de ces motifs pour aboutir à la qualification d'un phénomène. La Figure 3 présente l'ensemble des phénomènes inférés. Pour se faire, notre système se base sur l'utilisation de règles dites "d'inférences" qui permettent de générer de la connaissance automatiquement à partir de celle déjà présentes dans le système d'information. Il existe plusieurs outils dans la littérature permettant de fournir de telles règles. Dans ces travaux, nous utilisons le langage SPARQL 1.1² qui offre la possibilité d'interroger, créer, supprimer ou mettre à jour les informations présentes dans notre ontologie. L'algorithme 1 montre un exemple de règle permettant d'inférer un phénomène d'inondation en se basant sur un motif de séparation.

La *séparation* est le motif décrivant la division d'une entité spatiale avec une des parties qui conserve son identité (*Continuation*) et l'autre partie dont l'identité varie (*Dérivation*). Dans l'exemple, la partie de l'entité qui varie est passée de l'état de *zone urbaine* à l'état d'*étendue d'eau* et induit une réduction de la zone urbaine. En parallèle, la partie de l'entité qui n'a pas variée est également impliquée dans un motif d'*annexion* ce qui indique une *expansion urbaine* (il est à noter au passage que cette *expansion urbaine* se produit en empiétant sur la forêt connexe à la *zone urbaine* ce qui révèle un phénomène de *déforestation*). Enfin le dernier phénomène observable et qui peut être inféré sur le même principe de règle est un phénomène d'*intensification urbaine*. Ce phénomène peut être inféré lorsqu'un motif de *séparation* est observé sur une *zone urbaine* avec une partie se transformant en zone *industrielle et commerciale*.

2. <http://www.w3.org/TR/sparql11-query/>

Un modèle sémantique pour l'étude de la dynamique des territoires

```

insert
{
    ?p checksem:Inondation ?c
}
where
{
    ?p checksem:PartOfSeparation ?c .
    ?p a checksem:ArtificialSurfaces .
    ?c a checksem:Waterbodies .
}

```

Algorithme 1 : Règle SPARQL pour inférer un phénomène d'inondation

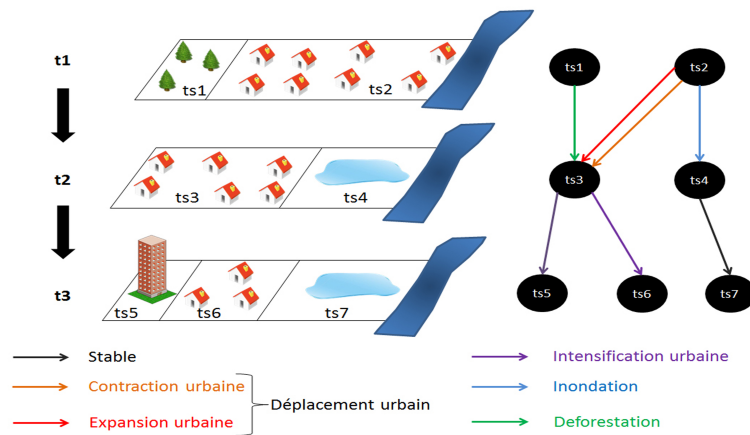


FIG. 3 – Liste des phénomènes inférés dans le modèle LC3

D'après les conclusions tirées de la précédente analyse, il est intéressant de noter qu'entre les instants T1 et T2, la zone urbaine est à la fois impliquée dans un phénomène d'*expansion* et de *contraction urbaine* ce qui peut sembler contradictoire de prime abord. En réalité la combinaison de ces deux phénomènes révèle un phénomène plus complexe qui est un *déplacement de la zone urbaine*. L'algorithme 2 présente la requête permettant de détecter ce type de phénomène complexe.

```

insert
{
    ?p checksem:UrbanMove ?c
}
where
{
    ?p checksem:hasAnnexation ?c .
    ?p checksem:hasSeparation ?c .
    ?p a checksem:ArtificialSurfaces .
    ?c a checksem:ArtificialSurfaces .
}

```

Algorithme 2 : Règle SPARQL pour inférer un phénomène de déplacement urbain

L'analyse fournie par notre modèle permet jusqu'ici de répondre à des questions telles que :

où, quoi, quand et comment a lieu le changement d'une entité spatio-temporelle ? Cependant, la question principale à laquelle un expert tente généralement de répondre est : *pourquoi ce changement a-t-il eu lieu ?*. Les causes à l'origine d'un phénomène sont souvent nombreuses et complexes, toutefois notre modèle offre des capacités accrues d'analyses en permettant de mettre en interaction différents phénomènes connexes. Par exemple, au travers du motif de *séparation* détecté entre T1 et T2, notre modèle établit une corrélation entre la *réduction du territoire urbain* et le phénomène d'*inondation*. De manière similaire, le motif d'*annexion* détecté entre T1 et T2 met en relation le phénomène d'*expansion urbaine* et le phénomène de *déforestation*. La *contraction* et l'*expansion urbaine* pouvant être assimilées à un *déplacement urbain*, il est alors aisé de conclure que la zone inondée a engendré un déplacement de la zone urbaine. Par conséquent, la forêt située en bordure de cette zone urbaine a subi un phénomène de *déforestation*.

6 Conclusion et travaux futurs

Fort du constat que les systèmes d'information géographiques actuels possèdent des capacités limitées pour représenter le contexte de l'environnement géospatial, nous avons proposé dans cet article un modèle basé sur les technologies du web sémantique capable de fournir des outils de raisonnement afin de faire émerger de nouvelles connaissances à partir de celles déjà présentes dans le système d'information. La structure sous forme de graphe permet d'autre part de travailler aisément sur la notion de relation. La relation de filiation permet de suivre les entités dans le temps ainsi que d'accroître la connaissance liée à l'évolution. Plusieurs couches d'expressivité ont été définies sur la base de cette relation. À terme, cette relation permet de révéler l'existence de phénomènes qui peuvent d'être mis en corrélation afin d'améliorer notre compréhension de l'évolution du territoire.

Notre modèle a été testé sur des jeux de données parcellaires uniquement. Nos futurs travaux s'orientent vers la prise en compte d'autres structures de données telles que les lignes ou les points. En reprenant l'exemple fourni à la fin de cet article, cela permettrait de mettre en corrélation les inondations avec la Garonne qui coule le long de la zone inondée. La cohabitation de ces différents types de données impose de définir des relations plus complexes ainsi que de nouveaux motifs génériques.

7 Remerciements

Ces travaux sont financés par 1) la Direction Générale de l'Armement, voir <http://www.defense.gouv.fr/dga/> 2) le Conseil régional de Bourgogne.

Références

- Armstrong, M. P. (1988). Temporality in spatial databases. In *Proceedings : GIS/LIS*, Volume 88, pp. 880–889.
- Artale, A. et E. Franconi (1998). A Temporal Description Logic for Reasoning About Actions and Plans. *Journal of Artificial Intelligence Research* 9(1), 463–506.

- Baader, F. et W. Nutt (2003). Basic description logics. *Description logic handbook*, 43–95.
- Claramunt, C. et M. Theriault (1996). Toward semantics for modelling spatio-temporal processes within gis. *Advances in GIS Research I*, 27–43.
- Del Mondo, G., M. A. Rodríguez, C. Claramunt, L. Bravo, et R. Thibaud (2013). Modeling Consistency of Spatio-temporal Graphs. *Data & Knowledge Engineering* 84, 59–80.
- Egenhofer, M. J. et K. K. Al-Taha (1992). Reasoning about gradual changes of topological relationships. *Theories and methods of spatio-temporal reasoning in geographic space*, 196–219.
- Harbelot, B., H. Arenas, et C. Cruz (2014). A semantic model to query spatial-temporal data. *Information Fusion and Geographic Information Systems (IF AND GIS 2013)*, 75–89.
- Hornsby, K. et M. J. Egenhofer (2000). Identity-based change : a foundation for spatio-temporal knowledge representation. *International Journal of Geographical Information Science* 14(3), 207–224.
- Jiang, J. et M. Worboys (2009). Event-based topology for dynamic planar areal objects. *International Journal of Geographical Information Science* 23(1), 33–60.
- Langran, G. et N. R. Chrisman (1988). A framework for temporal geographic information. *Cartographica : The International Journal for Geographic Information and Geovisualization* 25(3), 1–14.
- Muller, P. (2002). Topological spatio-temporal reasoning and representation. *Computational Intelligence* 18(3), 420–450.
- Peuquet, D. J. et N. Duan (1995). An event-based spatiotemporal data model (estdm) for temporal analysis of geographical data. *International journal of geographical information systems* 9(1), 7–24.
- Stell, J., G. Del Mondo, R. Thibaud, et C. Claramunt (2011). Spatio-temporal evolution as bigraph dynamics. *Spatial Information Theory*, 148–167.
- Worboys, M. F. (1994). A unified model for spatial and temporal information. *The Computer Journal* 37(1), 26–34.

Summary

It is necessary for land managers to be provided both an overview of existing knowledge and changes about the evolution of certain characteristics of the biosphere. Remote sensing tools monitore a large amount of land cover informations for studying dynamic processes. However, the volume of data sets requires new tools to identify patterns and extract knowledge. We propose a model enabling the knowledge discovery on landparcel data by analyzing dynamic phenomena using temporal, spatial and thematic data. The model is called Land Cover Change Continuum (LC3) and is based on Semantic Web technologies to enable a better representation about the context of the geospatial environment and provide analysis results close to those of experts ones via reasoning systems. Ultimately, this model can improve our understanding of spatio-temporal dynamics in land territories.